Center for Independent Experts (CIE) independent review report for the Bering Sea Tanner Crab Stock Assessment Review

Anders Nielsen

Executive Summary

This is an independent review of the Bering Sea Tanner Crab assessment. The review meeting focused more on the new TCSAM02 model framework instead of specific model configurations. The 'base model' that was described had convergence issues, which should be investigated further. Fixed likelihood weights, priors, and penalties were used, which may result in unreliable statistical inference (the confidence statements may be affected, but also the estimates, as the relative weight between information sources is shifted). The assessment framework TCSAM02 is well-designed with respect to the population dynamics, and it includes a wide range of options and is easily configurable. TCSAM02 is thoroughly validated against the previously used model framework TCSAM2013, which strengthens confidence in its implementation. A simulation study of the model would further validate the model, and it could also be used to show the correctness of the statistical properties of the model (unbiased estimates and correct coverage of the confidence statements). It is planned to implement a model for Tanner Crab in GMACS (Generalized Modeling for Alaskan Crab Stocks), and in doing so, it is important to carry over the lessons learned from TCSAM02 and from this review meeting.

Background

The Alaska Fisheries Science Center (AFSC) requested an independent review of the stock assessment and projection model framework 'TCSAM02' used to conduct the Bering Sea Tanner crab stock assessment. The review meeting took place in Seattle, Washington, from 31 July to 03 August 2017.

The model and projection framework was primarily presented by Dr. William T. Stockhausen and the meeting was chaired by Dr. Martin Dorn. The review panel consisted of (Drs. Cathy Dichmont, Australia; Norman Hall, Australia; and Anders Nielsen, Denmark). I thank everyone at the meeting for clear presentations and inspiring discussions. I am especially grateful for William's willingness to produce a great number of additional runs and outputs during the meeting.

This report documents the independent review of Anders Nielsen. The Statement of Work is appended to this report.

Description of the reviewer's role

This reviewer has independently read the assessment report, its appendices and all supplementary documents deemed necessary in preparation for the review, traveled and participated actively in the review meeting, identified key issues in the assessment, suggested guidance, fixed issue with model building software, and independently authored this review report.

Findings for each term of reference

To ensure that all terms of reference are covered and that comments are interpreted with reference to the correct terms, the terms are listed with corresponding reviewer comments following.

TOR 1: Statements assessing the strengths and weaknesses of the current Tanner crab stock assessment model with regard to population dynamics, fishery and survey components, likelihood components, and model evaluation.

The model framework TCSAM02 was presented in all details by Dr. William T. Stockhausen. The focus was mainly on the model framework and less on finding one specific model configuration for Tanner Crab. For the purpose of the review meeting (and this report), a model configuration candidate was named the 'base model', which was the model configuration 'T0A2' (p. 17 in the 'Tanner Crab Assessment Report for the May 2017 CPT Meeting').

The TCSAM02 model framework is a new and extended implementation of the previously used Tanner crab assessment model (TCSAM2013). The new implementation (TCSAM02) is more flexible, and allows and supports a wider range of model configurations with respect to model options, bounds, and priors. Further, it supports additional data types. TCSAM02 is implemented such that the different model configurations can be specified via input files, which is convenient because the user of the model framework generally does not need to recompile the code when testing different model options.

A huge effort has been undertaken to ensure that the new framework (TCSAM02) gives the same results as the old framework (TCSAM13) when configured equivalently. This strengthens confidence in both implementations.

It was reassuring to observe that Dr. William T. Stockhausen was able to quickly make changes to the model configuration and even to the model code to accommodate the requests from the reviewers. This flexibility is one of the benefits of having the model framework implemented by the assessment scientist compared to using a general model, where the assessment scientist is generally only able to manipulate the pre-specified model options.

The model framework (TCSAM02) is sufficiently documented with respect to its mathematical equations and likelihood components, but seems to be lacking a users guide. The different keywords used to invoke different model options are not described, which makes it impossible to operate the model framework without a fairly detailed knowledge of the source code behind the framework.

The model equations of TCSAM (both versions) are tailored especially to describe the population dynamics of Bering Sea Tanner Crab and the data available. It was stated that the long term goal for the Tanner Crab stock assessment is to implement a model in GMACS (Generalized Modeling for Alaskan Crab Stocks), which is another framework for size-based assessment modeling currently under development, and used for Red King Crab. The goal would be share code and quality control for a common framework to be used for most of the crab stocks. Currently, it is not possible to use GMACS for Tanner Crab, because options are lacking to describe the specific population dynamics of Tanner Crab. It was stated that work to a Tanner Crab model via GMACS could be expected to begin as early as the fall 2017.

It did seem like an unusual timing for a CIE review when plans are to move the model to a different framework (GMACS), but the strategy of the assessment group appears to be to use everything developed in TCSAM02 to expand GMACS. This way the model development, validation, and review process will remain relevant after the model has been moved to GMACS.

A strength of the model is that it closely describes the Bering Sea Tanner Crab population dynamics with respect to recruitment, growth, molting, and maturation. The additional natural mortality estimated in the isolated period from 1980-1984 to be able to account for the mismatch between catches and survey did appear a bit artificial, and missing information about the catch, in that period, seems to be an equally likely explanation.

With respect to the population dynamics part of the model another aspect to consider is to extend the model spatially, as there are indications of differences in the processes in the eastern and western parts of the population (size distributions and shell conditions). This is a huge undertaking but could be considered as a long-term research item.

The fishery and survey components of the model are similar to many commonly used assessment models. Fixed selection curves are assumed in time blocks before and after 1982. One concern is that some of the survey catchability parameters were hitting the pre-specified bounds. This is unfortunate because the bounds are generally not set based on knowledge about the catchabilities, but as a technical measure to force the parameter estimating routine to focus on the relevant area.

The likelihood function is intended to assign a measure of probability to the actual observed data for any combination model parameter values. The goal is to use this function to estimate the model parameters from the data.

The data likelihood for this model uses multinomial distributions to describe the length compositions, and normal or log-normal to describe total catches. This is standard practice in assessment models. In addition, TCSAM02 is able to use growth data from molt increments and here uses gamma distribution, which is consistent with the internal growth model.

Apart from the data driven contributions to the likelihood, there are contributions or modifications from prior distributions, overall weights of likelihood components, fixed variances, fixed sample sizes, and penalties for approaching bounds. These are all standard tools in any assessment practitioner's toolbox, as they can be useful to nudge a problematic model to converge to a sensible solution. However, it is a concern for the proposed base model that all these tools are used to a greater degree than desirable and often used in combination, such that their effect becomes impossible to figure out. The concern is that the estimated quantities of interest are influenced too much by the subjective penalties, and hence, not influenced enough by the observations.

Overall, the conclusion is that the assessment framework TCSAM02 is well developed, highly configurable, and has sufficient options to describe the Bering Sea Tanner Crab population and its data sources well. The work to ensure that the results match up with a similarly configured TCSAM2013 is an important quality control, which greatly strengthens confidence in the implementation. The base run(s) provided appeared to have convergence problems (more later), and the subjective penalties are a concern, so the search for a final set of settings should continue.

TOR 2: Statements assessing the strengths and weaknesses of the current Tanner crab stock projection model, with regard to methodology.

Forward projections are needed for the stock status determination and over fishing level (OFL) calculations. These follow Tier 3 for crab stocks as defined by the North Pacific Fishery Management Council.

A one year projection is conducted consistently with the model assuming average recruitment from 1982-present. An interesting detail is that this projection includes the expected catch of Tanner Crab in the snow crab fishery based on the F_{OFL} for snow crab (which then must be determined first).

As no reliable relationship between stock and recruitment can be identified, the MSY reference points are sensibly based on proxies.

The B_{MSY} proxy is defined as 35% of the equilibrium mature male biomass in the absence of fishing. This equilibrium state is found analytically and validated --- to all decimal points --- by simulation. The MSY proxy for fishing mortality is defined as the fishing mortality which results in a MMB at the proxy for B_{MSY}.

The projection model is largely consistent with the model used to describe the stock in the historic period. This is a strength because this model can potentially be validated by standard model diagnosis. The only potential concern is the treatment of recruitment. An average of estimated recruitments from 1982-present is used as a constant recruitment in all projected years. This is not consistent with the recruitments seen in the historic period, which fluctuated from year to year. Using a constant average value in place of fluctuating recruitments will result in a different representation uncertainty of the calculated equilibrium, which is likely not used anyway. However, it should also be validated that the equilibrium biomass found by assuming a constant mean value for recruitment in long term projections is corresponding to the equilibrium (mean or median) found by simulating recruitments consistently with the historic period. Realistic recruitments could be obtained by sampling with replacement from estimated recruitments from 1982-present, or by fitting a parametric distribution to the estimated recruitments (1982-present) and simulating from the fitted distribution. This validation should be fairly simple to conduct, because the code already exists to find the equilibrium by step-wise projection, so the only change is to use different recruitments in each step.

TOR 3: A review of the fishery dependent and independent data inputs to the stock assessment with regard to quality of information and appropriateness to the assessment.

The data used for Tanner crab are biological parameters directed and non-directed catches and data from an extensive survey.

The NMFS survey for Eastern Bering Sea Continental Shelf Bottom is a standard design based survey. It supplies size composition data, estimates of abundance biomasses, and estimated coefficients of variations. In the earliest years of the surveys, the area expanded, but in the last 30 years the survey area has been constant. The survey provides good coverage of the spatial area for the Tanner crab stock. The spatial area for Tanner crab appears well defined, which is also supported by the slope-study conducted in selected years. The trawl width and measurements techniques with respect to trawl width has also changed and improved over time. The survey is conducted yearly and mainly in the June-July period with 30-minute hauls. The catch is sorted and for a sub-sample carapace size, weight, chela height, and shell condition are determined.

Estimates of abundance biomass (and associated CV) are calculated from a stratified sampling procedure. A spatial model approach could potentially be used for expanding the area back in time and provide an objective way to fill in missing cells, and utilize double measurements and the extra 'hotspot' estimates. However, it is the impression that the current survey is of high quality and appropriate to use in the assessment.

The fishery dependent data consists of catches, discards, size compositions, and efforts for the directed fishery and for bycatches in the snow crab, red king crab, and groundfish fisheries.

Uncertainty for retained catch is trusted to be small, but in the beginning of the time series, there can be some uncertainty or bias with respect to snow versus Tanner crab, or in reported landings from non-US fleets. The discarded catch is likely more uncertain, but this is less of an issue because the model assumes a low handling mortality of ca. 32%.

The observer coverage is 30%, which means that observers are randomly allocated to 30% of boats and they are monitored for 100% of the trips. A small number of (randomly selected) pots are then processed by the observers.

Shell condition is used in the model, but the assignments to these categories is a bit subjective, so there could be an additional uncertainty or bias source here.

The overall impression of the fishery dependent data is that it is well sampled and appropriate to use in the assessment.

TOR 4: Recommendations for alternative approaches to evaluate model convergence and compare multiple models.

The model framework TCSAM02 includes all the needed components and options. The specific model runs presented did have some concerning issues with respect to convergence. This is not uncommon for assessment models of this size and complexity, but they should be resolved before the final model configuration is accepted.

The convergence issues with the current configurations are noticed from three different observations. The jitter diagnostics, convergence towards bounds.

The so-called jitter diagnostics (where the model is started in a number of different values in the hope that it -- with few or zero exceptions -- will converge to the same minimum) was being used in a creative and unconventional way. Instead of using it only to diagnose convergence it was used as part of the optimization procedure. A normal optimization routine (AD Model Builder's quasi-Newton) was started in each of the jittered starting values, and the one resulting in the lowest final likelihood was selected as the final converged model. This triggered a closer look at the individual jitter runs and each of their final points of convergence. For one model configuration, only two out of the 200 (from memory) jitters runs ended up in the same lowest value. For another model configuration, only about half of the jitters runs ended up in the same lowest value. The same lowest likelihood value did result in the same estimates of selected parameters of interest. This jitter analysis indicates that the likelihood surface for the selected model configuration has multiple local optimums. Selecting the best convergence point among the relative few jitter runs gives little confidence that the actual global optimum has been located. Notice that the parameter space for these models is more than 300 dimensional, so 200 initial points are not a lot.

All the model configurations investigated had some model parameters converging to their pre-specified bounds. A convergence towards a bound is only considered converged because the optimization routine (in AD Model Builder) adds a penalty, which is increasing as the bound is approached. If the penalty was removed, the negative log-likelihood surface would still be decreasing across the bound, so the real point of convergence could be outside the bound. The fact that the model converged towards the bounds for some parameters also indicates that convergence is to a point outside the jitter region, as the jitter initials were generated uniformly distributed on the interval from $L+0.2(U-L)$ to $U-0.2(U-L)$ (where U and L are upper and lower bounds respectively). Wider bounds were explored but only resulted in convergence to other bounds.

All model parameters were bounded via bound penalties. Bounds penalties are convenient to use in AD Model builder and often sufficient to nudge the model to converge to a point inside the given range (away from the bounds). A different technique for bounding model parameters is to use transformations. E.g. to bound a model parameter α between L and U an unbounded parameter $x \in (-\infty;\infty)$ is defined, but then $\alpha = \frac{1}{1+e^{-x}}(U-L)+L$ is used in the model. This technique does not use penalties, and hence the final likelihood value is only coming from the model (even if a parameter is close to a bound). Secondly, bounding by transformations is (in this reviewer's experience) less likely to get a parameter artificially trapped at a bound.

The profile likelihood was produced for the mature male biomass (MMB) in the last year, which is an important quantity derived from the model. The profile likelihood did not look anything like the standard Hessian based normal-approximation (Figure 1). The profile likelihood increased and decreased very sharply around the estimated point. This was first taken to indicate that the profile likelihood at the estimate was completely dominated by the bound penalty, and hence difficult to use for inference.

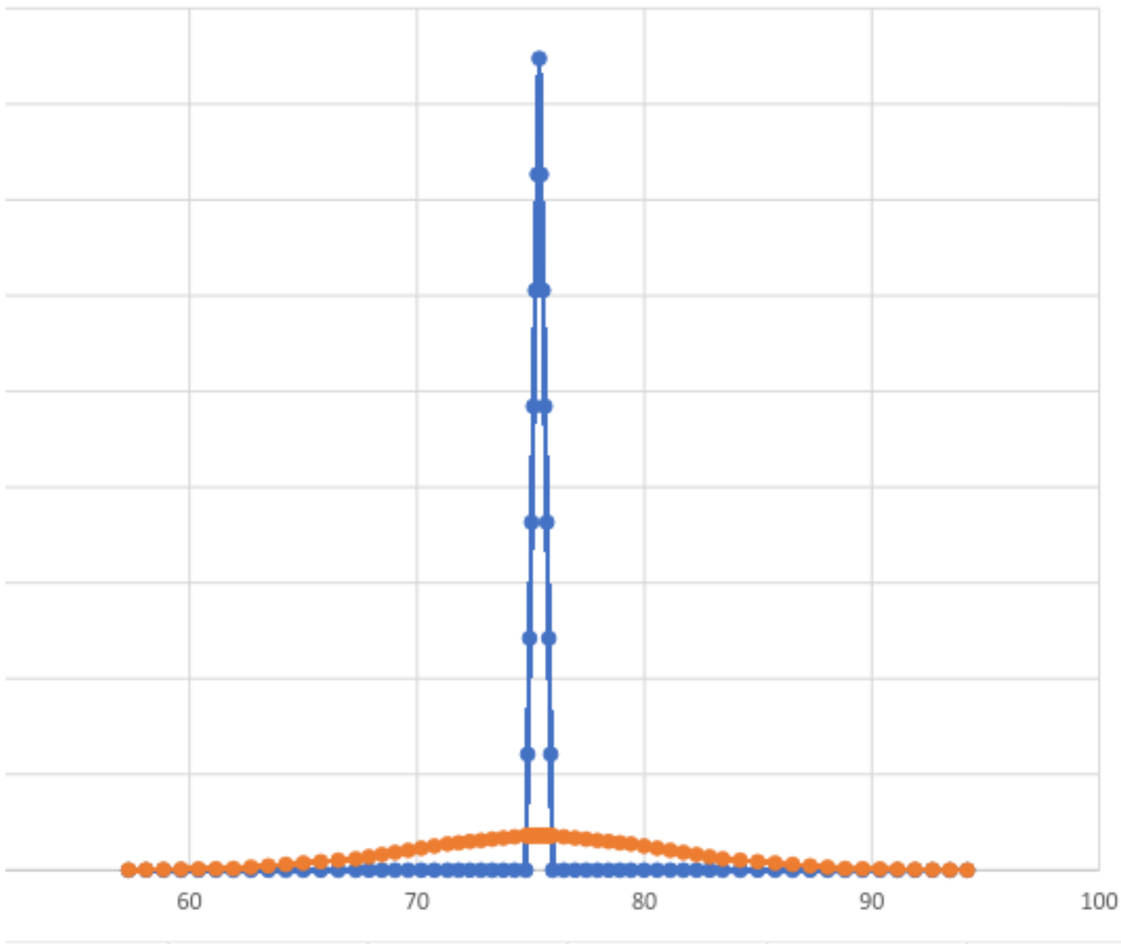

Figure 1: The profile likelihood of mature male biomass (blue) compared to the Hessian based normal approximation

However, after returning from the review meeting, the CIE reviewer Dr. Norman Hall remembered that he had seen a similar strange result before. Suspicion grew that this could be caused by a problem with some versions of the model building software AD Model Builder, which is used to implement TCSAM02. I tested examples that I knew had worked in the past and now they were failing in exactly the same way.

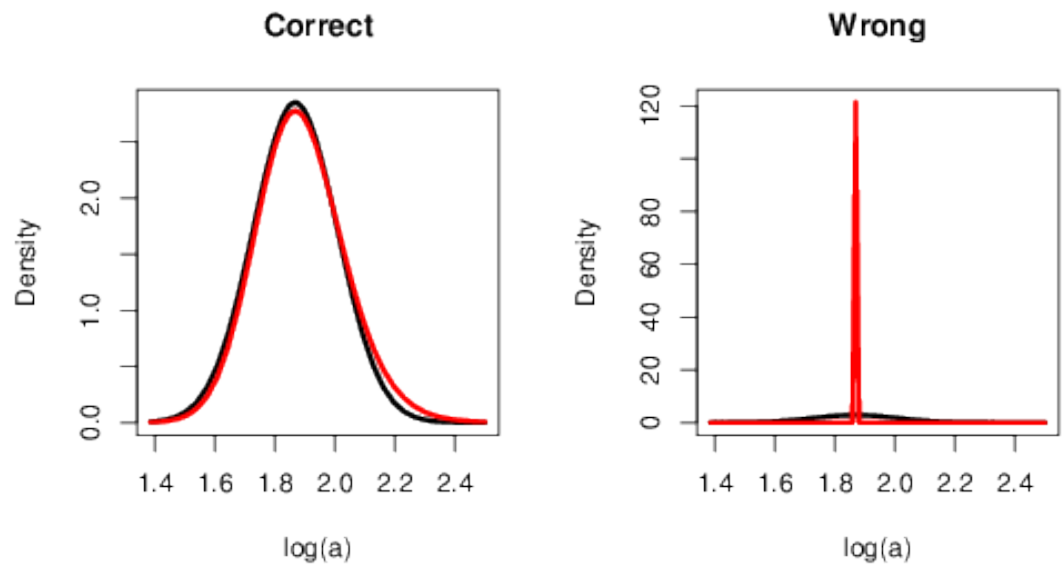

Figure 2: Likelihood profile (red line) compared to the Gaussian approximation (black line) in a simple case $log(a)$ in a Beverton-Holt model. The left panel shows the corrected version and the right panel when the bug is present. The code and data for this example is available in appendix 4.

Once I knew it had to be an error in AD Model builder, it was a straight forward (but fairly time-consuming) process to identify the code commit which caused the error. The code commit dated back to March 2014 and introduced a wrapper around the gradient calculation function. The wrapper worked for most things, but not for the profile likelihood calculation. The commit was rolled back and the profile likelihood was working again. To make sure that the fix to the profile likelihood did not break any other examples all the other built in examples were validated. To ensure that the profile likelihood is not broken again in the future, a test of the correctness of the profile likelihood was added, which will be tested after all future commits.

To get a problematic model to converge it is often necessary to simplify it first. Then when the model is sufficiently simplified (e.g. by fixing parameters) it can gradually be expanded up to the point where it is no longer converging. At this point, it is hopefully possible to figure out what is causing the lack of convergence.

It may be necessary to choose to fix a parameter, because it is confounded with some other parameter, and then illustrate the effect via a sensitivity study.

Apart from the data driven contributions to the likelihood the investigated model configurations contained contributions or modifications from prior distributions, fixed overall weights of likelihood components, fixed variances, fixed sample sizes, and penalties for approaching bounds. Many of these contributions are presumably added to get reasonable output or convergence, but at this point, the joint effect of all these is not transparent. It would likely be helpful to remove many of these constraints and fixate some model parameters instead, in order to figure out which parameters are really identifiable.

The jitter analysis is useful to detect a lack of convergence. Another tool is to conduct simulation studies. Simulate a few data sets from the model and verify that the estimated model parameters match up with the model parameters used to simulate the data sets. A simulation study is also an excellent tool to validate many other aspects of the model performance (coverage of confidence intervals, unbiased estimates, approximations, implementation, and much more).

When comparing assessment models, it is important to remember that the only thing we really have to compare them against is the actual observations. Quantities such as estimated biomasses and fishing mortalities are influenced by model settings, so comparing models against such quantities could lead to mistakes. Only after validating that the models sufficiently describe data is it sensible to compare e.g. retrospective performance with respect to estimated quantities. When comparing two models, it is useful to look at model validation e.g. residuals and Q-Q-plots. If one model is clearly better at describing the data then that is a good sign.

An often overlooked way to compare models is to compare predictions. A primary use for assessment models is to provide predictions, so an obvious way to compare them is to compare how well they are able to predict the *observations* one year ahead. The procedure is simply to optimize the model to the first N years of data and then predict the observation for the year $N + 1$ by whatever mechanism the model would use to provide predictions (some models are more consistent than others).

Finally, the statistical properties of the -- at this point remaining models -- should be investigated, which includes things like bias, confidence coverage, and retrospective patterns.

TOR 5: Recommendations for integrating BSFRF surveys into the assessment.

The Bering Sea Fisheries Research Foundation (BSFRF) survey is a set supplementary tows, which are paired (side-by-side) with some of the tows from the standard NMFS survey. The supplementary tows are conducted with a fine mesh (Nephrops trawl). The selectivity is assumed to be 100% for the considered size classes in the BSFRF survey.

This data can be incorporated in the assessment model in many ways e.g.: 1) use the study only to estimate the catchability of the standard NMFS survey. 2) used both the NMFS and the BSFRF to produce a joint improved survey index. 3) include both survey data series into the assessment model. In all cases, it seems important to take account of the paired nature of the samples. Paired samples can be expected to be correlated, as they are taken at (almost) the same location. If paired samples are mistakenly treated as independent samples in the model, then the final uncertainty estimates will not be correct (too small), but the estimates themselves may also differ because the relative weighting of the different samples will be incorrect.

E.g. if the goal is purely to inform about the catchability of the standard NMFS survey, then a model could be something along the lines of the following. Let catches from both surveys be denoted c_i. Here i is a simple index of all single catches-at-size, so corresponding to each catch is a station number $stat_i$, a size class $size_i$, a survey type sur_i, and a year $year_i$. Assume that the catches follow a Poisson distribution $c_i \sim pois(\lambda_i)$, where:

$$\log\lambda_i = \log q(sur_i) + 1_{\{sur_i=NMFS\}} \cdot \log S(size_i) + \alpha(size_i, year_i) + U(size_i, year_i, stat_i)$$

In this formulation, there is a catchability for each of the two survey types, a selection function S for the survey type 'NMFS' (the BSFRF is assumed to have full selectivity for all included sizes), a relative log abundance size α, and finally a random effect U, which is intended to describe the between station variation. Assume $U \sim N(0, \sigma_s^2)$. Notice that two paired observations will be described by the same U, and hence be correlated.

This is just a raw outline. In real application, the assumptions must be verified, and it could possibly be anticipated that catches from neighboring length classes at the same station could be further correlated, but this can also be accounted for via random effects.

Models like these can sometimes be handled via R-packages for general linear mixed effects models, but more generally by using random effects in AD Model Builder (Fournier et al. 2012), or in TMB (Kristensen et al. 2016). Similar models have been used in Benoit & Cadigan (2014).

The supplementary BSFRF tows cover the first 5 minutes of the corresponding NMFS tows. It is possible that the selectivity of the NMFS trawl changes as the trawl fills up, and hence that the first 5 minutes are not representative for the average selectivity of the NMFS trawl.

TOR 6: Recommendations for alternative assessment/projection model configurations.

The TCSAM02 assessment framework contains all the right components and can be easily configured and adjusted by the assessment lead Dr. William T. Stockhausen.

The likelihood part of the model configuration could be improved by reducing the use of 1) bounds penalties, 2) weights, 3) fixed variances, 4) deviance penalties, 5) and prior distributions. The combination of all of these things makes the inference about quantities less transparent. Some suggestions could be:

1) Replacing bounded model parameters with unbounded model parameters transformed into the proper domain (see under TOR 4).

2) removing the overall weights for the different likelihood components --- it is often confounded with the variance parameters anyway. These likelihood weights are a quick way of testing influence of different parts of the model, so the option should not be removed, but in the final model formulation, it is simpler to explain the model if these weights are all removed.

3) It is preferable if the model can be formulated, such that it is self-weighting (variances estimated). It is not always possible, so when it is not fixing variances (or equivalently

sample sizes) can be necessary. In that case, it is important to show the effect by presenting sensitivity studies.

4) Deviance penalties are fixed parameters. Thus, if used, the effects of the deviance penalties should be investigated by sensitivity analysis. An alternative to using a deviance parameter vector with fixed deviance variance is to use random effects. A vector of parameters (random effects) can be declared and assumed to follow a normal distribution with mean zero and a variance, which is a parameter to be estimated. This is a very useful approach to describe many quantities which are time-varying (see e.g. Nielsen and Berg 2014).

5) The use of prior distributions can be a shortcut to use additional information without including the data providing the information. The priors are then setup via estimates and distribution of estimates from an independent study. When priors are instead used more subjectively to nudge a problematic model convergence, or simply to express prior expert knowledge, then it is important to clearly demonstrate the effects of the prior information. A useful way is to plot the prior density together with the posterior density, because it shows how much of the final estimate is due to observations and how much is simply prior belief. Sensitivities for important outputs are also useful.

If the convergence is still problematic after changing the way bounds are handled, then the model should be configured to fix parameters until stable convergence can be achieved. The assessment lead has informed the reviewers that he is currently considering reformulating the survey selectivity, such that it is possible to fix the size-at-99%-selected to a size near the upper range of the sizes observed in the survey (125 mm CW for females, 175 mm CW for males). This sounds promising, because the survey catchabilities were among some of the parameters converging to a bound, so they could reasonably be suspected to be confounded.

It is recommended to study the effect of sampling from recent recruitment estimates (1982-present) instead if using the average for the long-term projections (to equilibrium). See description under TOR 2.

TOR 7: Recommendations for research that would reduce the uncertainty associated with key parameters assumed or estimated in the assessment.

It is important to understand the selectivity of the survey gear, and the additional BSFRF survey is available, so including that survey appears to be an important improvement at a relatively low effort.

If the BSFRF survey proves to be an important addition with respect to estimating the selectivity of the NMFS, then it should also be studied if it can improve the predictions. The BSFRF survey could be used to give an index of the smaller size groups, and hence earlier states. Its usefulness could be tested by prediction of catches and NMFS indices. If it improved the predictions, it should be considered if the BSFRF survey should be continued for more years.

Because AD Model Builder was producing the wrong profile likelihood, it is still unknown what the true profile likelihood looks like. The profile likelihood, and especially the profile likelihood for each likelihood component (data source) can show what parts of data are contradicting each other, which could be causing uncertainty in estimated key parameters.

An alternative to the stratified sampling procedure currently used to calculate biomass indices could be replaced with a spatial model approach. A spatial model approach could potentially be used to expand the area back in time and provide an objective way to fill in missing cells, and utilize double measurements and the extra 'hotspot' estimates.

Data on movement will be important if the model at a later point will be made more spatially explicit, so it is important to collect data to prepare for that. The movement study with acoustic pingers seems promising (see Pedersen and Weng 2013 for modeling of such data).

TOR 8: Suggested priorities for future improvements to the stock assessment/projection model.

Some fairly large population differences were seen between eastern and western parts of the stock area in particular with respect to the size compositions. This could suggest that a more spatially explicit model (possibly including migration) could perform better. It is however not a small task to formulate, implement, and configure these models. Spatial models also result in more complex models (with more model parameters) and it is difficult to predict if the advantage of a more flexible spatial model is big enough to justify the added complexity.

Random effects (a.k.a. latent variables or state-space models) can be a useful way to express flexible models with few model parameters. They can be useful for time-varying effects (as an alternative to penalized deviance vectors without the subjective assignment of a penalty variance), but even more so within spatially varying population models. Great advances have been made in tools to efficiently handle models with random effects (both in AD Model builder, but even more so in Template Model Builder (TMB)).

References

Benoit, H.P., and Cadigan, N. 2014. Model-based estimation of commercial-sized snow crab (Chionoecetes opilio) abundance in the southern Gulf of St. Lawrence, 1980-2012, using data from two bottom trawl surveys. DFO Can. Sci. Advis. Sec. Res. Doc. 2013/114. v + 46 p.

Fournier, D.A, Skaug, H.J, Ancheta, J, Ianelli, J, Magnusson, A, and Maunder, M.N, Nielsen, A, and Sibert. J. 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optimization Methods and Software 27 (2), 233-249.

Kristensen, K, Nielsen, A, Berg, C.W, Skaug, H.J, Bell, B. 2016. TMB: Automatic differentiation and laplace approximation Journal of Statistical Software 70 (5), 1-21.

Nielsen, A and Berg, C.W. 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fisheries Research 158, 96-101.

Pedersen, M.W, Weng, K.C. 2013. Estimating individual animal movement from observation networks. Methods in Ecology and Evolution 4 (10), 920-929.

Appendix 1: List of documents.

Statement of Work. National Oceanic and Atmospheric Administration (NOAA). National Marine Fisheries Service (NMFS). Center for Independent Experts (CIE) Program. External Independent Peer Review. Bering Sea Tanner Crab Stock Assessment Review. [SoW peer review_TannerCrab.pdf]

Tentative Agenda (2017-06-27). Bering Sea Tanner Crab Stock Assessment Review. [TannerCrab_CIEReviewAgenda.20170627.pdf]

Office of Management and Budget. Memorandum M-05-03. Final Information Quality Bulletin for Peer Review. [OMB_Peer_Review_Bulletin_m05-03.pdf]

Assessment model runs [AssessmentModelRuns folder]

Base model [BaseModel folder]

ModelRun folder: Folder containing input data and model output for base model.

Miscellaneous data files for alternative model runs, together with R code to prepare model output for use with R

Base model plus growth data [BaseModel+GrowthData folder]

ModelRun folder: Folder containing input data and model output for base model.

Miscellaneous data files for alternative model runs, together with R code to prepare model output for use with R

Final model [FinalModel folder]

ModelRun folder: Folder containing input data and model output for base model.

Miscellaneous data files for alternative model runs, together with R code to prepare model output for use with R

Model comparisons [ModelComparisons folder]

R Markdown code and output for model comparisons and differences for B0, B1, and B2.

R packages [R_Packages folder]

Compressed files containing R code for utilities to process model output.

Notes on model runs [NotesOnModelRuns.docx]

Files containing executable code, i.e. runTCSAM02.pin.bat, runTCSAM02.pin.sh, tcsam02.exe, and tcsam02.osx)

Documents

Assessment documents

Report to May 2017 CPT Meeting [201705ReportToCPT folder]

Stockhausen, W. T. (2017). Tanner Crab Assessment Report for the May 2017 CPT Meeting. Alaska Fisheries Science Center, April 2017.

Appendix A: Corrected Retained Catch Size Frequencies in the Directed Tanner crab Fisheries.

Appendix B: Tanner crab growth (molt increment) data.

Appendix C: Tanner Crab Bycatch in the Groundfish Fisheries.

Appendix D1: Model Comparisons for TCSAM2013 Models B0, B1, B2, B3, B4, B5, B6.

Appendix D2: Model Differences for B1, B2, B3, B4, B5 and B6 vs. B0.

Appendix E: Model Differences between T13B6 and T02A.

Appendix F1a: Model Comparisons for T02A vs AG0.

Appendix F1b: Model Differences for T02A vs AG0.

Appendix F2a: Model Comparisons for AG1 vs AG0.

Appendix F2b: Model Differences for AG1 vs AG0.

Appendix F3a: Model Comparisons for AG1, AG2a, AG2b and AG3.

Appendix F3b: Model Differences for AG1, AG2a, AG2b, and AG3.

Appendix G1: Model Comparisons for TCSAM02 Models AG1, AG1a, AG1b and AG1d.

Appendix G2: Model Differences for TCSAM02 Models AG1a, AG1b, and AG1d vs. AG1.

Appendix H1: Model Comparisons for TCSAM02 Models AG1 and AG1c.

Appendix H2: Model Differences for TCAM02 Models AG1c vs AG1.

Appendix I1: Model Comparisons for TCSAM02 Models AG1 and AG1e.

Appendix J2: Model Differences for TCSAM02 Models AG1 vs AG1e

Appendix J1: Model Comparisons for TCAM02 Models AG3, AG3a, AG3b, and AG4.

Appendix J2: Model Differences for AG3a, AG3b and AG4 vs. AG3.

Appendix K1: Model Comparisons for TCSAM02 Models B1, AG4, and AG1c.

Appendix K2: Model Differences for TCSAM02 Models B1, AG4, and AG1c.

NPFSC BSAI Crab SAFE. (2016). 2016 Stock Assessment and Fishery Evaluation Report for the King and Tanner Crab Fisheries in the Bering Sea and Aleutian Islands [2016 Introduction Chapter Crab SAFE.pdf]

North Pacific Fishery Management Council's Crab Plan Team. (2016). Crab Plan Team Report. October 2016. [CrabPlanTeamReport_2016-10.pdf]

Punt, A., and Kinzey, D. (prep. by) Report of the Alaska Crab Stock Assessment Workshop, 13-14 May 2009. [SAFE_Guidelines.pdf]

Stockhausen, W. T. (2016).2016 Stock Assessment and Fishery Evaluation Report for the Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions. September 2016 Plan Team draft. [TannerCrab201609_SAFEReport.pdf]

Stockhausen, W. T. (2017). 2017 Crab Modeling Workshop Report for Tanner Crab. [TannerCrab201701_ModelingWorkshop.pdf]

Assessment model description

Anon. (no date) TCSAM02: The Tanner Crab Stock Assessment Model, version 2. [TCSAM02Description.201707.CIEReview.pdf]

Biology

Donaldson, W. E., Cooney, R. T., and Hilsinger, J. R. (1981). Growth, age and size at maturity of Tanner crab, *Chionoecetes bairdi* M. J. Rathbun, in the northern Gulf of Alaska (Decapoda, Brachyura). *Crustaceana*, 40: 286–302. [DonaldsonEtAl1981_Crustaceana.pdf]

Jadamec, L. S., Donaldson, W. E., and Cullenberg, P. (1999). *Biological Field Techniques for Chionoecetes Crabs*. University of Alaska Sea Grant College Program; Fairbanks, Alaska. [AK-SG-99-02PDF-Biological Field Techniqu.pdf]

Livingston, P. A. (1989). Interannual trends in Pacific Cod, *Gadus macrocephalus*, predation on three commercially important crab species in the eastern Bering Sea. *Fishery Bulletin*, **87**: 807–827. [Livingston1989_FishBull.PredationMortality.pdf]

Paul, A. J., and Paul, J. M. (1995). Molting of functionally mature male *Chionoecetes bairdi rathbun* (decapoda: majidae) and changes in carapace and chela measurements. Journal of Crustacean Biology, **15**: 686–692. [Paul_Paul1995b_JCrustBiol.Growth.pdf]

Siddeek, M. S. M., Zheng, J., Morado, J. F., Kruse, G. H., and Bechtol, W. B. (2010). Effect of bitter crab disease on rebuilding in Alaska Tanner crab stocks. *ICES Journal of Marine Science*, **67**: 2027–2032. [Sideek2010_ICESJMS.BitterCrabDiseaseMortalityForTannerCrab.pdf]

Somerton, D. A. (1981). Regional variation in the size of maturity of two species of tanner crab (*Chionoecetes bairdi* and *C. opilio*) in the eastern Bering Sea, and its use in defining management subareas. *Can. J. Fish. Aquat. Sci*., **38**: 163- 174. [Somerton1981_CJFAS.Maturity.pdf]

Stoner, A. W., Rose, C. S., Munk, J. E., Hammond, C., and Davis, M. W. (2008). An assessment of discard mortality for two Alaskan crab species, Tanner crab (*Chionoecetes bairdi*) and snow crab (*C. opilio*) based on reflex impairment. *Fish. Bull*., **106**: 337–347. [StonerEtAl_FishBull2008_RAMP.HandlingMortality.pdf]

Urban, D., and Hart, D. 1999 Biology of the tanner crab *Chionoecetes bairdi* in Alaska: a report to the Alaska Board of Fisheries. Kodiak, Alaska: Alaska Dept. of Fish and Game, Division of Commercial Fisheries. Regional Information Report No. 4K99-22. [Urban_Hart_RIR.4K.1999.22.pdf]

Zheng, J. (2008). Temporal changes in size at maturity and their implications for fisheries management for eastern Bering Sea Tanner crab. *J. Northw. Alt. Fish. Sci*., **41**: 137–149. [Zheng2008_JNorthwAtlFishSci.Maturity.pdf]

Zheng, J., and Kruse, G. H. (1998). Stock–recruitment relationships for Bristol Bay Tanner crab. *Alaska Fish. Res. Bull.*, **5**: 116–130. [ZhengAndKruse1998_AFRB.StockRecruitRelationships.pdf]

Fisheries

Leon, J. M., Shaishnikoff, J., Nichols, E., and Westphal, M. (2017). *Annual Management Report for Shellfish Fisheries in the Bering Sea–Aleutian Islands Management Area, 2015/16*. Alaska Department of Fish and Game. Fishery Management Report No. 17-10. [ADFG.FisheryManagementReport17-10.pdf]

Alaska Department of Fish and Game. (no date). 2015–2017 King and Tanner Crab Commercial Fishing Regulations. Alaska Department of Fish and Game, Anchorage, Alaska. [ADFG.Regulations_KingTannerCrab-2015-2017.pdf]

Stockhausen, W. T. (2017). *A brief history of the Tanner crab fishery*. Unpublished. [TannerCrabFisheryHistory.pdf]

Management

Staff of the Alaska Fisheries Science Center, Seattle and Kodiak. (2008). *Final Environmental Assessment for Amendment 24 to the Fishery Management Plan for Bering Sea/Aleutian Islands King and Tanner crabs to revise overfishing definitions.* [NPFMC.FinalEAforAmendment24.pdf]

Anon. (2011). *Fishery Management Plan for Bering Sea/Aleutian Islands King and Tanner Crabs.* North Pacific Fishery Management Council, Anchorage, Alaska. [NPFMC.FMP_Crab2011.pdf]

Surveys

NMFS

Daly, B. J., Armistead, C. E., and Foy, R. J. (2016). The 2016 Eastern Bering Sea Continental Shelf Bottom Trawl Survey: Results for Commercial Crab Species. NOAA Technical Memorandum NMFS-AFSC-327. [NOAA-TM-AFSC-327_2016_EBS_ShelfSurvey.pdf]

Somerton, D. A., and Otto, R. S. (1999). Net efficiency of a survey trawl for snow crab, *Chionoecetes opilio*, and Tanner crab, *C. bairdi*. Fish. Bull. **97**:617–625. [Somerton_Otto_FishBull1999.Selectivity.pdf]

BSFRF

Stockhausen, W. T. (2017). BSFRF Side-by-Side Survey Results. [BSFRFSurveys.pdf]

Appendix 2: Statement of work for Anders Nielsen.

Statement of Work

National Oceanic and Atmospheric Administration (NOAA)

National Marine Fisheries Service (NMFS)

Center for Independent Experts (CIE) Program

External Independent Peer Review

Bering Sea Tanner Crab Stock Assessment Review

Background

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available (BSIA). NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards. (http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf).

Further information on the CIE program may be obtained from www.ciereviews.org.

Scope

The Alaska Fisheries Science Center (AFSC) Resource Ecology and Fishery Management (REFM) Division requests an independent review of the stock assessment/projection model used to conduct the Bering Sea Tanner crab stock assessment. The model is a size-based integrated assessment model and has been under continuous development since being approved for use by the North Pacific Fisheries Management Council (NPFMC) in 2012. It is anticipated that the North Pacific Fisheries Management Council's Crab Plan Team (CPT) and Science and Statistical Committee (SSC) will approve a change in the TCSAM (Tanner Crab Stock Assessment) code used for the assessment from "TCSAM2013", the code used for the 2013-2016 assessments, to "TCSAM02", a new modeling framework that provides a much more flexible environment than TCSAM2013 for defining alternative models based on a set of model configuration files, as well as fitting new data types not incorporated in TCSAM2013: molt increment (growth) and male chela height (maturity) data. TCSAM02 also calculates the OFL and associated quantities directly within a model run, and thus retains full model uncertainty when using MCMC, whereas using TCSAM2013 the OFL is calculated in a separate projection model and incorporates uncertainty only in recruitment and end-year mature biomass. This review will encompass the TCSAM02 stock assessment/projection model structure and assumptions on which it is based, as well as the life history, fishery, and survey data incorporated in the model. It will also address alternatives for incorporating several industry-funded surveys into the assessment. The Terms of Reference (TORs) for the requested peer review are described in more detail in Annex 2.

Requirements

NMFS requires three (3) CIE reviewers with the necessary qualifications to complete an impartial and independent peer review in accordance with the tasks and TORs (Annex 2) described in the Statement of Work (SOW) herein. The CIE reviewers shall have expertise in conducting stock assessments for fisheries management and be thoroughly familiar with various subject areas involved in stock assessment, including population dynamics, size-structured models, harvest strategies, survey methodology, and the AD Model Builder programming language to complete the tasks of the scientific peer-review described herein. Familiarity with invertebrate stock assessment, knowledge of crab life history and biology, and harvest strategy development is desirable.

Tasks for Reviewers

- Review the following background materials and reports prior to the review meeting:

    1. *Stockhausen, W. 2017. May 2017 Tanner Crab Stock Assessment Activities Report. In prep. [For review:*

    2. *Stram, D. et al. 2016. Introduction Chapter. In: 2016 Stock Assessment and Fishery Evaluation Report for the King and Tanner Crab Fisheries in the Bering Sea and Aleutian Islands. North Pacific Fisheries Management Council, Anchorage, AK. http://npfmc.legistar.com/gateway.aspx?M=F&ID=2f46b828-51ca-4a45-95bb-cddae2ed8f1d.pdf. [Review the "Stock Status Definitions" and "Status Determination Criteria" for background on the NPFMC's crab stock status criteria and OFL determination]*

    3. *Stockhausen, W. 2016. 2016 Stock Assessment and Fishery Evaluation Report for the Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions. In: 2016 Stock Assessment and Fishery Evaluation Report for the King and Tanner Crab Fisheries in the Bering Sea and Aleutian Islands. North Pacific Fisheries Management Council, Anchorage, AK. http://npfmc.legistar.com/gateway.aspx?M=F&ID=0e48278f-160e-426b-972e-f4736e7c8726.pdf. [The last stock assessment, based on the TCSAM2013 model code.]*

    4. *DALY, B. J., C. E. ARMISTEAD, and R. J. FOY. 2016. The 2016 eastern Bering Sea continental shelf bottom trawl survey: Results for commercial crab species, 167 p. NTIS No. PB2016-104795. [Report on the 2016 NMFS annual eastern Bering Sea shelf summer crab/groundfish trawl survey.]*

    5. *A document (TBD) describing the Gmacs assessment framework.*

    6. *A document (TBD) describing the BSFRF surveys*

- Attend and participate in the panel review meeting:
    - The meeting will consist of presentations by NOAA and other scientists, stock assessment authors and others to facilitate the review, to provide any additional information required by the reviewers, and to answer any questions from reviewers.
- After the review meeting, reviewers shall conduct an independent peer review in accordance with the requirements specified in this SOW, OMB guidelines, and TORs, in adherence with the required formatting and content guidelines; reviewers are not required to reach a consensus.
- Each reviewer may assist the Chair of the meeting with contributions to the summary report, if required by the TORs.
- Deliver their reports to the Government according to the specified milestone dates.

Foreign National Security Clearance

When reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for reviewers who are non-US citizens. For this reason, the reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/ and http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-registration-system.html. The contractor is required to use all appropriate methods to safeguard Personally Identifiable Information (PII).

Place of Performance

Each CIE reviewer shall participate in, and conduct an independent peer review during, the panel review meeting at the Alaska Fisheries Science Center (AFSC) in Seattle, Washington. Pre- and post-review performance shall be conducted at the contractor's facilities.

Period of Performance

The period of performance shall be from the time of award through September 15, 2017. Each reviewer's duties shall not exceed 14 days to complete all required tasks.

Schedule of Milestones and Deliverables: The contractor shall complete the tasks and deliverables in accordance with the following schedule.

Within two weeks of award	**Contractor selects and confirms reviewers**
No later than 17 July 2017	Contractor provides the pre-review documents to the reviewers
31 July – 3 August 2017	Panel review meeting
17 August 2017	Contractor receives draft reports
7 September 2017	Contractor submits final reports to the Government

Applicable Performance Standards

The acceptance of the contract deliverables shall be based on three performance standards:

(1) The reports shall be completed in accordance with the required formatting and content as described in Annex 1; (2) The reports shall address each TOR as specified in Annex 2; (3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

Travel

All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (http://www.gsa.gov/portal/content/104790). International travel is authorized for this contract. Travel is not to exceed $14,000.

Restricted or Limited Use of Data

The contractors may be required to sign and adhere to a non-disclosure agreement.

NMFS Project Contact

William Stockhausen

william.stockhausen@noaa.gov

National Marine Fisheries Service

7600 Sand Point Way, NE, Bldg. 4,

Seattle, WA 98115-6349

Phone: (206) 526-4241

Annex 1: Peer Review Report Requirements

1. The report must be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether or not the science reviewed is the best scientific information available.

2. The report must contain a background section, description of the individual reviewers' roles in the review activities, summary of findings for each TOR in which the weaknesses and strengths are described, and conclusions and recommendations in accordance with the TORs.

   a. Reviewers must describe in their own words the review activities completed during the panel review meeting, including a brief summary of findings, of the science, conclusions, and recommendations.

   b. Reviewers should discuss their independent views on each TOR even if these were consistent with those of other panelists, but especially where there were divergent views.

   c. Reviewers should elaborate on any points raised in the summary report that they believe might require further clarification.

   d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.

   e. The report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The report shall represent the peer review of each TOR, and shall not simply repeat the contents of the summary report.

3. The report shall include the following appendices:

Appendix 1: Bibliography of materials provided for review

Appendix 2: A copy of this Statement of Work

Appendix 3: Panel membership or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference for the Peer Review

Bering Sea Tanner Crab Stock Assessment Review

The report generated by the consultant should include:

1. Statements assessing the strengths and weaknesses of the current Tanner crab stock *assessment* model with regard to population dynamics, fishery and survey components, likelihood components, and model evaluation.
2. Statements assessing the strengths and weaknesses of the current Tanner crab stock *projection* model, with regard to methodology.
3. A review of the fishery dependent and independent data inputs to the stock assessment with regard to quality of information and appropriateness to the assessment.
4. Recommendations for alternative approaches to evaluate model convergence and compare multiple models.
5. Recommendations for integrating BSFRF surveys into the assessment.
6. Recommendations for alternative assessment/projection model configurations.
7. Recommendations for research that would reduce the uncertainty associated with key parameters assumed or estimated in the assessment.
8. Suggested priorities for future improvements to the stock assessment/projection model.

Annex 3: Tentative Agenda

Bering Sea Tanner Crab Stock Assessment Review

NOAA Alaska Fisheries Science Center

7600 Sand Point Way NE

Seattle, WA 98115

August 2017

Monday

09:00Welcome and Introductions

09:15Role of chair and reviewers, terms of reference

09:30Overview (fishery, catch levels, bycatch, surveys)

10:30Biology (growth, natural mortality, maturity curves, mating, molting frequency)

12:00Lunch

13:00Survey methodology

14:30Fishery history and current operation

15:30Harvest control rules and overfishing definition

17:00Evening break

Tuesday

09:00Stock assessment and projection model

12:00Lunch

13:00Stock assessment and projection model (continued)

17:00Evening break

Wednesday

9:00Current research studies

growth, fecundity and egg production

BSFRF side-by-side surveys and other research

12:00Lunch

1300Strategies for integrating BSFRF surveys into assessment

14:00Gmacs

17:00Evening break

Thursday

9:00Reviewer discussions with assessment author.

Review of requested model runs if required.

Appendix 3: List of participants.

Participants at the CIE review of the Tanner crab stock assessment, July 31-August 3, 2017

Name, Affiliation

Martin Dorn, AFSC, Meeting chair

William Stockhausen, AFSC, Lead assessment author

Jack Turnock, AFSC

Anne Hollowed, AFSC

Jeff Napp, AFSC

Ben Daly, Alaska Department of Fish and Game, Kodiak

Scott Goodman, Bering Sea Fisheries Research Foundation

Gary Stauffer, Bering Sea Fisheries Research Foundation

Alathea Letaw, University of Washington

Remote:

Robert Foy, AFSC Kodiak

Miranda Westphal, Alaska Department of Fish and Game, Dutch Harbor

CIE reviewers:

Cathy Dichmont, Cathy Dichmont Consulting, Australia

Anders Nielsen, Technical University of Denmark, Denmark

Norman Hall, Murdoch University, Western Australia

AFSC: Alaska Fisheries Science Center

Appendix 4a: Code for example in figure 2 (to test if ADMB is working correctly).

```
DATA_SECTION
  init_int nR
  init_int nC
  init_matrix obs(1,nR,1,nC)
  vector ssb(1,nR)
  !! ssb=column(obs,1);
  vector logR(1,nR)
  !! logR=column(obs,2);
PARAMETER_SECTION
  init_number loga;
  init_number logb;
  init_number logSigma;
  sdreport_number sigmaSq;
  vector pred(1,nR);
  likeprof_number loga_pl       // NOTICE
  likeprof_number logb_pl       // NOTICE
  objective_function_value nll;
PRELIMINARY_CALCS_SECTION       //!
  loga_pl.set_stepnumber(50);   //!
  loga_pl.set_stepsize(0.1);    //!
  logb_pl.set_stepnumber(50);   //!
  logb_pl.set_stepsize(0.1);    //!
PROCEDURE_SECTION
  loga_pl=loga;                 //NOTICE
  logb_pl=logb;                 //NOTICE
  sigmaSq=exp(2.0*logSigma);
  pred=loga+log(ssb)-log(1+exp(logb)*ssb);
  nll=0.5*(nR*log(2*M_PI*sigmaSq)+sum(square(logR-pred))/sigmaSq);
```

Appendix 4a: Data for example in figure 2 (to test if ADMB is working correctly).

```
46 2
136650 12.62
150340 13.28
188240 13.42
210350 12.91
239390 13.21
255950 13.99
249430 13.41
259650 13.52
268300 13.23
242970 13.46
213260 13.52
227870 13.59
211000 13.95
181600 12.52
156780 13.18
157780 12.73
159430 13.31
174690 13.68
188650 13.07
189440 12.7
154600 13.39
129290 13.49
120930 13.12
113150 12.53
105380 12.89
98576 12.59
93093 12.47
77977 13.35
70771 12.99
67379 12.82
64748 12.75
69805 12.32
80186 12.81
78796 12.96
75325 12.81
61501 12.77
56104 12.29
47022 12.98
39408 12.38
41458 12.69
36851 12.23
32716 11.83
31164 12.04
27362 12.01
33775 11.66
39558 11.83
```